Note

Alternative implementations of Monte Carlo EM algorithms for likelihood inferences

Louis Alberto GARCÍA-CORTÉS[a], Daniel SORENSEN[b],[*]

[a] Departamento de Genética, Universidad de Zaragoza,
Calle Miguel Servet 177, Zaragoza, 50013, Spain
[b] Section of Biometrical Genetics, Department of Animal Breeding and Genetics,
Danish Institute of Agricultural Sciences, PB 50, 8830 Tjele, Denmark

Abstract – Two methods of computing Monte Carlo estimators of variance components using restricted maximum likelihood *via* the expectation-maximisation algorithm are reviewed. A third approach is suggested and the performance of the methods is compared using simulated data.

restricted maximum likelihood / Markov chain Monte Carlo / EM algorithm / Monte Carlo variance / variance components

1. INTRODUCTION

The expectation-maximisation (EM) algorithm [1] to obtain restricted maximum likelihood (REML) estimators of variance components [7] is widely used. The expectation part of the algorithm can be demanding in highly dimensional problems because it requires the inverse of a matrix of the order of the number of location parameters of the model. In animal breeding this can be of the order of hundred of thousands or millions.

Guo and Thompson [3] proposed a Markov chain Monte Carlo (MCMC) approximation to the computation of these expectations. This is useful because in principle it allows to analyse larger data sets but at the expense of introducing Monte Carlo noise. Thompson [9] suggested a modification to the algorithm which reduces this noise.

The purpose of this note is to review briefly these two approaches and to suggest a third one which can be computationally competitive to the Thompson estimator.

* Correspondence and reprints
E-mail: sorensen@inet.uni2.dk

2. THE MODEL AND THE EM-REML EQUATIONS

The sampling model for the data is assumed to be

$$\mathbf{y}|\mathbf{b}, \mathbf{s},\sigma_e^2 \sim N\left(\mathbf{Xb} + \mathbf{Zs}, \mathbf{I}\sigma_e^2\right)$$

where $\mathbf{y}$ is the vector of data of length n, $\mathbf{X}$ and $\mathbf{Z}$ are incidence matrices, $\mathbf{b}$ is a vector of fixed effects of length p, $\mathbf{s}$ is a vector of random sire effects of length q, $\mathbf{I}$ is the identity matrix and $\mathbf{I}\sigma_e^2$ is the variance associated with the vector of residuals $\mathbf{e}$. Sire effects are assumed to follow the Gaussian process:

$$\mathbf{s}|\sigma_s^2 \sim N\left(\mathbf{0}, \mathbf{I}\sigma_s^2\right)$$

where σ_s^2 is the variance component due to sires.

Implementation of restricted maximum likelihood with the EM algorithm requires setting the conditional expectations (given $\mathbf{y}$) of the natural sufficient statistics for the model of the complete data $(\mathbf{y}, \mathbf{s})$ equal to their unconditional expectations. Let $\boldsymbol{\theta} = \left(\sigma_s^2, \sigma_e^2\right)$. In the case of the present model, the unconditional expectations of the sufficient statistics are $E\left(\mathbf{s}'\mathbf{s}|\boldsymbol{\theta}\right) = q\sigma_s^2$ and $E\left(\mathbf{e}'\mathbf{e}|\boldsymbol{\theta}\right) = n\sigma_e^2$. The conditional expectations require computation of

$$E\left(\mathbf{s}'\mathbf{s}|\mathbf{y},\hat{\boldsymbol{\theta}}\right) = \hat{\mathbf{s}}'\hat{\mathbf{s}} + \mathrm{tr}\left[\mathrm{Var}\left(\mathbf{s}|\mathbf{y},\hat{\boldsymbol{\theta}}\right)\right] = \hat{\mathbf{s}}'\hat{\mathbf{s}} + \sum_{i=1}^{q} \mathrm{Var}\left(s_i|\mathbf{y},\hat{\boldsymbol{\theta}}\right) \qquad (1)$$

and

$$E\left(\mathbf{e}'\mathbf{e}|\mathbf{y},\hat{\boldsymbol{\theta}}\right) = \left(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{s}}\right)'\left(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}} - \mathbf{Z}\hat{\mathbf{s}}\right) + \mathrm{tr}\left[\mathrm{Var}\left(\mathbf{e}|\mathbf{y},\hat{\boldsymbol{\theta}}\right)\right], \qquad (2)$$

where $\hat{\boldsymbol{\theta}}$ is the value of $\boldsymbol{\theta}$ at the current EM iterate, and $\hat{\mathbf{b}}$ and $\hat{\mathbf{s}}$ are the expected values of $\left[\mathbf{b}, \mathbf{s}|\mathbf{y},\hat{\boldsymbol{\theta}}\right]$ which can be obtained as the solution to Henderson's mixed model equations [5]:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{I}\hat{k} \end{bmatrix}\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{s}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}. \qquad (3)$$

In (3), $\hat{k} = \hat{\sigma}_e^2/\hat{\sigma}_s^2$. Throughout this note, to make the notation less cumbersome, a parameter x with a "hat" on top, $\hat{x}$, will refer to the value of the parameter at the current EM iterate.

3. THE GUO AND THOMPSON ESTIMATOR

A MC estimate of the expectation in (1) requires realizations from the distribution $\left(\mathbf{s}|\mathbf{y},\hat{\boldsymbol{\theta}}\right)$. A restricted maximum likelihood implementation of the approach proposed by Guo and Thompson [3] involves drawing successively from

$$p\left(s_i|\mathbf{s}_{-i}, \mathbf{b},\hat{\boldsymbol{\theta}}, \mathbf{y}\right)$$

and from

$$p\left(b_j|\mathbf{b}_{-j}, \mathbf{s},\hat{\boldsymbol{\theta}}, \mathbf{y}\right),$$

where $i = 1, \ldots, q; j = 1, \ldots, p$, using the Gibbs sampler [2]. In this notation x_i is a scalar, and the vector $\mathbf{x}_{-i}$ is equal to the vector $\mathbf{x} = \{x_i\}$ with x_i excluded. With T realisations from $\left(\mathbf{s}|\mathbf{y},\hat{\boldsymbol{\theta}}\right)$, the MC estimate of the variance component at the current EM iterate is

$$\hat{\sigma}_s^2 = \frac{1}{qT} \sum_{i=1}^{T} \left(\mathbf{s}^{(j)\prime}\mathbf{s}^{(j)}\right) \tag{4}$$

where T is the number of rounds of Gibbs samples used within the current EM iterate. After several cycles, once convergence has been reached, the converged values are averaged to obtain the final MCEM REML estimator. Guo and Thompson [3] provide a detailed description of the algorithm and a useful overview can be found in [8].

4. THE THOMPSON ESTIMATOR

Consider the decomposition of $\mathrm{Var}\left(s_i|\mathbf{y},\hat{\boldsymbol{\theta}}\right)$ in (1):

$$\mathrm{Var}\left(s_i|\mathbf{y},\hat{\boldsymbol{\theta}}\right) = E\left[\mathrm{Var}\left(s_i|\mathbf{s}_{-i}, \mathbf{b}, \mathbf{y},\hat{\boldsymbol{\theta}}\right)\right] + \mathrm{Var}\left[E\left(s_i|\mathbf{s}_{-i}, \mathbf{b}, \mathbf{y},\hat{\boldsymbol{\theta}}\right)\right]$$

$$= \left(\mathbf{z}_i'\mathbf{z}_i + \hat{k}\right)^{-1} \hat{\sigma}_e^2 + \mathrm{Var}\left[E\left(s_i|\mathbf{s}_{-i}, \mathbf{b}, \mathbf{y},\hat{\boldsymbol{\theta}}\right)\right] \tag{5}$$

where $\mathbf{z}_i$ is the ith row of the incidence matrix $\mathbf{Z}$. In (5), only the second term has MC variance. This term is given by

$$\mathrm{Var}\left[E\left(s_i|\mathbf{s}_{-i}, \mathbf{b}, \mathbf{y},\hat{\boldsymbol{\theta}}\right)\right] = E_{\mathbf{s}_{-i},\mathbf{b}|\mathbf{y},\hat{\theta}}\left[E\left(s_i|\mathbf{s}_{-i}, \mathbf{b}, \mathbf{y},\hat{\boldsymbol{\theta}}\right)^2\right]$$

$$- \left\{E_{\mathbf{s}_{-i},\mathbf{b}|\mathbf{y},\hat{\theta}}\left[E\left(s_i|\mathbf{s}_{-i}, \mathbf{b}, \mathbf{y},\hat{\boldsymbol{\theta}}\right)\right]\right\}^2$$

$$= E_{\mathbf{s}_{-i},\mathbf{b}|\mathbf{y},\hat{\theta}}\left[\tilde{s}_i^2\right] - \hat{s}_i^2 \tag{6}$$

where $\tilde{s}_i = E\left(s_i | \mathbf{s}_{-i}, \mathbf{b}, \mathbf{y}, \hat{\boldsymbol{\theta}}\right)$ and $\hat{s}_i = E\left(s_i | \mathbf{y}, \hat{\boldsymbol{\theta}}\right)$. Using (6) and (5) in (1) yields:

$$E\left(\mathbf{s}'\mathbf{s} | \mathbf{y}, \hat{\boldsymbol{\theta}}\right) = \sum_{i=1}^{q} \left\{ \left(\mathbf{z}'_i \mathbf{z}_i + \hat{k}\right)^{-1} \hat{\sigma}_e^2 + E_{\mathbf{s}_{-i}, \mathbf{b} | \mathbf{y}, \hat{\theta}} \left[\tilde{s}_i^2\right] \right\}. \tag{7}$$

The MC estimator of (7) is given by $\sum_{i=1}^{q} \left\{ \left(\mathbf{z}'_i \mathbf{z}_i + \hat{k}\right)^{-1} \hat{\sigma}_e^2 + \frac{1}{T} \sum_{j=1}^{T} \left[\tilde{s}_i^{(j)2}\right] \right\}$, where $\tilde{s}_i^{(j)}$ is the expected value over the fully conditional distribution $\left[s_i | \mathbf{s}_{-i}, \mathbf{b}, \mathbf{y}, \hat{\boldsymbol{\theta}}\right]$ at the jth Gibbs round $(j = 1, \ldots, T)$. The MC estimate of the variance component at the current EM iterate is now equal to

$$\hat{\sigma}_s^2 = \frac{1}{q} \sum_{i=1}^{q} \left\{ \left(\mathbf{z}'_i \mathbf{z}_i + \hat{k}\right)^{-1} \hat{\sigma}_e^2 + \frac{1}{T} \sum_{j=1}^{T} \left(\tilde{s}_i^{(j)2}\right) \right\}$$

$$= \frac{1}{q} \sum_{i=1}^{q} \left(\mathbf{z}'_i \mathbf{z}_i + \hat{k}\right)^{-1} \hat{\sigma}_e^2 + \frac{1}{qT} \sum_{j=1}^{T} \left(\tilde{\mathbf{s}}^{(j)'} \tilde{\mathbf{s}}^{(j)}\right). \tag{8}$$

This expression is equivalent to equation (4) in [9].

5. AN ALTERNATIVE ESTIMATOR

Consider the distribution of $\left[\mathbf{s} | \hat{\boldsymbol{\theta}}, \mathbf{y} = \mathbf{0}\right]$, which is normal, with mean zero and variance equal to the variance of $\left[\mathbf{s} | \hat{\boldsymbol{\theta}}, \mathbf{y}\right]$. That is:

$$\left[\mathbf{s} | \hat{\boldsymbol{\theta}}, \mathbf{y} = \mathbf{0}\right] \sim N\left(\mathbf{0}, \text{Var}\left(\mathbf{s} | \hat{\boldsymbol{\theta}}, \mathbf{y} = \mathbf{0}\right)\right). \tag{9}$$

The term $\text{Var}\left(\mathbf{s} | \hat{\boldsymbol{\theta}}, \mathbf{y} = \mathbf{0}\right)$ corresponds to the lower diagonal block of the inverse of the coefficient matrix of (3) at the current EM iterate. Then equation (1) can be written:

$$E\left(\mathbf{s}'\mathbf{s} | \mathbf{y}, \hat{\boldsymbol{\theta}}\right) = \hat{\mathbf{s}}'\hat{\mathbf{s}} + \sum_{i=1}^{q} \text{Var}\left(s_i | \mathbf{y} = \mathbf{0}, \hat{\boldsymbol{\theta}}\right). \tag{10}$$

Decomposing the second term in (10), using (5) and (6), and noting that

$$E_{\mathbf{s}_{-i}, \mathbf{b} | \mathbf{y}, \hat{\theta}} \left[E\left(s_i | \mathbf{s}_{-i}, \mathbf{b}, \mathbf{y} = \mathbf{0}, \hat{\boldsymbol{\theta}}\right)\right] = 0$$

yields:

$$\mathrm{Var}\left(s_i|\mathbf{y}=\mathbf{0},\hat{\boldsymbol{\theta}}\right) = \left(\mathbf{z}_i'\mathbf{z}_i+\hat{k}\right)^{-1}\hat{\sigma}_e^2 + \mathrm{Var}\left[E\left(s_i|\mathbf{s}_{-i},\mathbf{b},\mathbf{y}=\mathbf{0},\hat{\boldsymbol{\theta}}\right)\right]$$

$$= \left(\mathbf{z}_i'\mathbf{z}_i+\hat{k}\right)^{-1}\hat{\sigma}_e^2 + E_{\mathbf{s}_{-i},\mathbf{b}|\mathbf{y}=\mathbf{0},\hat{\theta}}\left[E\left(s_i|\mathbf{s}_{-i},\mathbf{b},\mathbf{y}=\mathbf{0},\hat{\boldsymbol{\theta}}\right)^2\right]$$

$$= \left(\mathbf{z}_i'\mathbf{z}_i+\hat{k}\right)^{-1}\hat{\sigma}_e^2 + E_{\mathbf{s}_{-i},\mathbf{b}|\mathbf{y}=\mathbf{0},\hat{\theta}}\left[\tilde{s}_{0,i}^2\right]$$

where $\tilde{s}_{0,i} = E\left(s_i|\mathbf{s}_{-i},\mathbf{b},\mathbf{y}=\mathbf{0},\hat{\boldsymbol{\theta}}\right)$. Therefore

$$E\left(\mathbf{s}'\mathbf{s}|\mathbf{y},\hat{\boldsymbol{\theta}}\right) = \hat{\mathbf{s}}'\hat{\mathbf{s}} + \sum_{i=1}^{q}\left\{\left(\mathbf{z}_i'\mathbf{z}_i+\hat{k}\right)^{-1}\hat{\sigma}_e^2 + E_{\mathbf{s}_{-i},\mathbf{b}|\mathbf{y},\hat{\theta}}\left[\tilde{s}_{0,i}^2\right]\right\}. \qquad (11)$$

The MC estimator of (11) is given by

$$\hat{\mathbf{s}}'\hat{\mathbf{s}} + \sum_{i=1}^{q}\left\{\left(\mathbf{z}_i'\mathbf{z}_i+\hat{k}\right)^{-1}\hat{\sigma}_e^2 + \frac{1}{T}\sum_{j=1}^{T}\left[\tilde{s}_{0,i}^{(j)2}\right]\right\}$$

where $\tilde{s}_{0,i}^{(j)}$ is the expected value over the fully conditional distribution $\left[s_i|\mathbf{s}_{-i},\right.$ $\mathbf{b},\mathbf{y}=\mathbf{0},\hat{\boldsymbol{\theta}}\right]$ at the jth Gibbs round. The MC estimate of the variance component at the current EM iterate is now equal to

$$\hat{\sigma}_s^2 = \frac{1}{q}\left[\hat{\mathbf{s}}'\hat{\mathbf{s}} + \sum_{i=1}^{q}\left\{\left(\mathbf{z}_i'\mathbf{z}_i+\hat{k}\right)^{-1}\hat{\sigma}_e^2 + \frac{1}{T}\sum_{j=1}^{T}\left(\tilde{s}_{0,i}^{(j)2}\right)\right\}\right]. \qquad (12)$$

6. COMPARISON OF MONTE CARLO VARIANCES

The smaller MC variance of (8) relative to (4) is explained *via* the decomposition of the variance of $\left[s_i|\mathbf{y},\hat{\boldsymbol{\theta}}\right]$ in (5): only the second term is subject to MC variance.

In order to compare the MC variance of (12) and (8) note that the ith element of $\tilde{\mathbf{s}}^{(j)}$ in (8), $\tilde{s}_i^{(j)}$, can be written as

$$\tilde{s}_i^{(j)} = \hat{s}_i + \tilde{s}_{0,i}^{(j)}. \qquad (13)$$

Inserting (13) in (8) yields

$$\hat{\sigma}_s^2 = \frac{1}{q}\left\{\hat{\mathbf{s}}'\hat{\mathbf{s}} + \sum_{i=1}^{q}\left[\left(\mathbf{z}_i'\mathbf{z}_i+\hat{k}\right)^{-1}\hat{\sigma}_e^2 + \frac{1}{T}\sum_{j=1}^{T}\left(2\hat{s}_i\tilde{s}_{0,i}^{(j)} + \tilde{s}_{0,i}^{(j)2}\right)\right]\right\} \qquad (14)$$

which shows that (8) has an extra term relative to (12) which contributes with extra MC variance.

Table I. Restricted maximum likelihood estimators of sire variance based on expression (4) − method 1, expression (8) − method 2, and (12) − method 3. The REML of σ_s^2 and the MC variance of σ_s^2 is based on 1 000 Monte Carlo replicates.

Estimator	True h^2	REML of σ_s^2	MC variance of σ_s^2
1	0.1	5.69	0.0707
2	0.1	5.71	0.0007
3	0.1	5.71	0.0001
1	0.3	11.08	0.0995
2	0.3	11.10	0.0025
3	0.3	11.10	0.0003
1	0.5	19.97	0.0949
2	0.5	19.98	0.0047
3	0.5	19.98	0.0008

6.1. An example with simulated data

The performance of the three methods is illustrated using three simulated data sets with heritabilities equal to 10%, 30% and 50%. In each data set, one thousand offspring records distributed in 100 herds were simulated from 100 unrelated sires. The figures in Table I show the performance of the three methods in terms of their MC variances, which were computed empirically based on 1 000 replicates.

The figures in Table I show clearly the ranking of the methods in terms of their MC variances.

Table II shows a comparison of (8) and (12) in terms of computing time (these two only are shown because the time taken to run (4) and (8) is almost identical). The length of the MC chain for updating σ_s^2 at each EM iterate (T) was increased for each method, until the same MC variance of 0.0005 was obtained, and the time taken was recorded.

Estimator (12) requires knowledge of $E\left(\mathbf{s}|\mathbf{y},\hat{\boldsymbol{\theta}}\right)$ at each EM iterate. Therefore, it takes longer per EM iteration than (8). However, the proposed method is still more efficient than that based on (8) since it compensates by requiring a shorter MC chain length (T) for updating σ_s^2 at each EM iterate. In general, the relative efficiency of the methods will be dependent on the model and data structure.

7. EXTENSION TO CORRELATED RANDOM EFFECTS

MCEM provides an advantage in models where the E step is difficult to compute, and this can be the case in models with many correlated random

Table II. CPU-time in seconds (per complete EM replicate) taken for estimators (8) and (12) to achieve an MC variance equal to 0.0005. (Based on 100 replicates).

Estimator [a]	True h^2	T [b]	CPU-time (s)
2	0.1	60	9.12
3	0.1	6	7.96
2	0.3	190	29.28
3	0.3	26	16.15
2	0.5	400	60.74
3	0.5	90	41.95

[a] Same symbol as in Table I.
[b] Length of MC chain for updating σ_s^2 at each EM iterate.

effects. An example is the additive genetic model, where each phenotypic observation has an associated additive genetic value. Consider the model

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}^*\mathbf{a} + \mathbf{e}$$

where $\mathbf{y}$ is the data vector of length n, $\mathbf{X}$ and $\mathbf{Z}^*$ are known incidence matrices, $\mathbf{b}$ is a vector of fixed effects of length p and $\mathbf{a}$ is the vector of additive genetic values of length q. With this model, usually $q > n$. The sampling model for the data is

$$\mathbf{y}|\mathbf{b}, \mathbf{a}, \sigma_e^2 \sim N\left(\mathbf{Xb} + \mathbf{Z}^*\mathbf{a}, \mathbf{I}\sigma_e^2\right)$$

where σ_e^2 is the residual variance. Vector $\mathbf{a}$ is assumed to be multivariate normally distributed as follows:

$$\mathbf{a}|\mathbf{A}, \sigma_a^2 \sim N\left(\mathbf{0}, \mathbf{A}\sigma_a^2\right)$$

where $\mathbf{A}$ is the $q \times q$ additive genetic covariance matrix and σ_a^2 is the additive genetic variance.

Consider the decomposition of $\mathbf{A}$ as in [6]:

$$\mathbf{A} = \mathbf{FF}'.$$

Define the transformation

$$\boldsymbol{\varphi} = \mathbf{F}^{-1}\mathbf{a}.$$

It follows that

$$\boldsymbol{\varphi}|\sigma_a^2 \sim N\left(\mathbf{0}, \mathbf{I}\sigma_a^2\right).$$

Let $\boldsymbol{\theta} = \left(\sigma_a^2, \sigma_e^2\right)$. Clearly, conditionally on $\mathbf{y}$ and $\hat{\boldsymbol{\theta}}$ the distribution of $\left(\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}\right)$ and of $\left(\boldsymbol{\varphi}'\boldsymbol{\varphi}\right)$ have the same expectation; that is:

$$E\left(\mathbf{a}'\mathbf{A}^{-1}\mathbf{a}|\mathbf{y},\hat{\boldsymbol{\theta}}\right) = E\left(\boldsymbol{\varphi}'\boldsymbol{\varphi}|\mathbf{y},\hat{\boldsymbol{\theta}}\right)$$

$$= \hat{\boldsymbol{\varphi}}'\hat{\boldsymbol{\varphi}} + \sum_{i=1}^{q} \text{Var}\left(\varphi_i|\mathbf{y},\hat{\boldsymbol{\theta}}\right)$$

where $\hat{\boldsymbol{\varphi}}$ satisfies:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{I}\hat{k} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\boldsymbol{\varphi}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}. \tag{15}$$

In (15), $\mathbf{Z} = \mathbf{Z}^*\mathbf{F}$. We label the coefficient matrix in (15), $\mathbf{W} = \left\{w_{ij}\right\}; i,j = 1,\ldots,p+q$.

The MCEM restricted maximum likelihood estimator of σ_a^2 at the current EM iterate, equivalent to (4) is now

$$\hat{\sigma}_a^2 = \frac{1}{qT} \sum_{j=1}^{T} \left(\boldsymbol{\varphi}'^{(j)}\boldsymbol{\varphi}^{(j)}\right)$$

where $\boldsymbol{\varphi}^{(j)}$ is the vector of Gibbs samples at the jth round, whose elements are drawn from $p\left(\varphi_i|\mathbf{y},\hat{\boldsymbol{\theta}}\right)$, $i = 1,\ldots,q$. Using the same manipulations as before that led to (8) and (12), it is easy to show that the estimators equivalent to (8) and (12) are

$$\frac{1}{q} \sum_{i=1}^{q} \left(w_{i+p,i+p}^{-1}\right) \hat{\sigma}_e^2 + \frac{1}{qT} \sum_{j=1}^{T} \left(\tilde{\boldsymbol{\varphi}}'^{(j)}\tilde{\boldsymbol{\varphi}}^{(j)}\right) \tag{16}$$

and

$$\frac{1}{q} \left[\hat{\boldsymbol{\varphi}}'\hat{\boldsymbol{\varphi}} + \sum_{i=1}^{q} w_{i+p,i+p}^{-1}\hat{\sigma}_e^2 + \frac{1}{T} \sum_{j=1}^{T} \left(\tilde{\boldsymbol{\varphi}}_0'^{(j)}\tilde{\boldsymbol{\varphi}}_0^{(j)}\right) \right] \tag{17}$$

respectively. In (16), $\tilde{\boldsymbol{\varphi}}^{(j)}$ is the vector of Gibbs samples at the jth round with elements $\tilde{\varphi}_i = E\left(\varphi_i|\boldsymbol{\varphi}_{-i}, \mathbf{b}, \mathbf{y},\hat{\boldsymbol{\theta}}\right)$. Similarly in (17), $\tilde{\boldsymbol{\varphi}}_0^{(j)}$ is the vector of Gibbs samples at the jth round with elements $\tilde{\varphi}_{0,i} = E\left(\varphi_i|\boldsymbol{\varphi}_{-i}, \mathbf{b}, \mathbf{y} = \mathbf{0},\hat{\boldsymbol{\theta}}\right)$. In both expressions, $w_{i+p,i+p}^{-1}$ is the inverse of the diagonal element of row/column $i+p$ of design matrix $\mathbf{W}$.

8. DISCUSSION

Markov chain Monte Carlo (MCMC) methods are having an enormous impact in the implementation of complex hierarchical statistical models. In the present paper we discuss three MCMC-based EM algorithms to compute Monte Carlo estimators of variance components using restricted maximum likelihood. Another application along the same lines can be found in [4], which shows how to use the Gibbs sampler to obtain elements of the inverse of the coefficient matrix which features in equations (1) and (2).

The performance of the methods was compared in terms of their Monte Carlo variances and in terms of length of computing time to achieve a given Monte Carlo variance. The different behaviour of the methods is disclosed in expressions (4), (14) and (12). The method based on (8) divides the overall sum of squares involved in (4) into two terms, one of which has no Monte Carlo noise. In our method, further partitioning is achieved which includes a term which is not subject to Monte Carlo noise. However, this is done at the expense of requiring a solution to a linear system of equations. When tested with simulated data, the proposed method performed better than the other two. The data and model used induced a simple correlation structure. The relative performance of the proposed method may well be different with models that generate a more complicated correlation structure.

Efficient implementation is likely to require a fair amount of experimentation. For example, the solution to the linear system in each round within an EM iterate need only be approximate and can be used as starting values for the next iteration. Similarly, the number of iterates within each round (variable T) can be tuned with the total number of cycles required to achieve convergence. One possibility which we have not explored is to set $T = 1$ and to let the system run until convergence is reached.

REFERENCES

[1] Dempster A.P., Laird N.M., Rubin D.B., Maximum likelihood from incomplete data with EM algorithm, J. Roy. Stat. Soc. B 39 (1977) 1–38.

[2] Gelfand A.E., Smith A.F.M., Sampling based approaches to calculating marginal densities, J. Am. Stat. Assoc. 85 (1990) 398–409.

[3] Guo S.W., Thompson E.A., Monte Carlo estimation of variance component models for large complex pedigrees, IMA J. Math. Appl. Med. Biol. 8 (1991) 171–189.

[4] Harville D.A., Use of the Gibbs sampler to invert large, possibly sparse, positive definite matrices, Linear Algebra and its Applications 289 (1999) 203–224.

[5] Henderson C.R., Sire evaluation and genetic trends, in: Proc. Animal Breeding and Genetics Symposium in honor of Dr. Jay L. Lush, Blacksburg, VA, August 1973, American Society of Animal Science, Champaign, IL, pp. 10–41.

[6] Henderson C.R., A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values, Biometrics 32 (1976) 69–83.

[7] Patterson H.D., Thompson R., Recovery of inter–block information when block sizes are unequal, Biometrika 58 (1971) 545–554.

[8] Tanner M.A., Tools for statistical inference, Springer–Verlag, New York, 1996.

[9] Thompson R., Integrating best linear unbiased prediction and maximum likelihood estimation, in: Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Guelph, 7–12 August 1994, University of Guelph, Vol. XVIII, pp. 337–340.